Noise Texture Deviation: A Measure for Quantifying Artifacts in Computed Tomography Images With Iterative Reconstructions

Morsbach, Fabian ; Desbiolles, Lotus ; Raupach, Rainer ; Leschka, Sebastian ; Schmidt, Bernhard ; Alkadhi, Hatem

Abstract: OBJECTIVES The aims of this study were to introduce the measure noise texture deviation as quantitative parameter for evaluating iterative reconstruction (IR)-specific artifacts in computed tomography (CT) images and to test whether IR-specific artifacts, quantified through this measure, are reduced in advanced modeled IR (ADMIRE) as compared with sinogram-affirmed IR (SAFIRE) images of the liver ex vivo and in patients with hypodense liver lesions. MATERIALS AND METHODS In the ex vivo study part, an abdominal phantom was used. In the institutional review board-approved in vivo study part, 40 consecutive patients (mean age, 63 years) with hypodense liver lesions undergoing abdominal CT in the portal-venous phase were included. Images were reconstructed with filtered back projection, with the second-generation IR algorithm SAFIRE and with the third-generation IR algorithm ADMIRE. Noise power spectra and noise texture deviation were calculated in the phantom; image noise was measured in the phantom and in patients. Two blinded readers evaluated all image data regarding IR-specific artifacts (plastic-like, blotchy appearance); patient data were evaluated regarding conspicuity and confidence for detecting hypodense liver lesions. RESULTS Image noise was significantly reduced at increasing IR levels ($P < 0.001$) with both algorithms, with no significant differences between corresponding strength levels of SAFIRE and ADMIRE (all, $P > 0.05$). Noise power spectra were similar at corresponding strength levels of SAFIRE and ADMIRE (all, $P > 0.05$). Noise texture deviation in ADMIRE was reduced compared with corresponding strength levels of SAFIRE (all, $P < 0.001$) and strongly correlated with subjective IR-specific artifacts ($r = 0.88$, $P < 0.001$). Iterative reconstruction-specific artifacts were significantly reduced in ADMIRE compared with that in SAFIRE images at strength levels 3 or greater, both ex vivo and in vivo (all, $P < 0.001$). There were no significant differences in the readers' ratings of lesion conspicuity and lesion confidence in detecting hypodense liver lesions between SAFIRE and ADMIRE ($P > 0.05$). Only lesion conspicuity was superior with SAFIRE and ADMIRE compared with filtered back projection (all, $P < 0.001$). CONCLUSIONS Noise texture deviation is a quantitative measure reflecting IR-specific artifacts and is reduced in CT images with ADMIRE compared with SAFIRE.

DOI: https://doi.org/10.1097/RLI.0000000000000312

Morsbach, Fabian; Desbiolles, Lotus; Raupach, Rainer; Leschka, Sebastian; Schmidt, Bernhard; Alkadhi, Hatem (2017). Noise Texture Deviation: A Measure for Quantifying Artifacts in Computed Tomography Images With Iterative Reconstructions. Investigative Radiology, 52(2):87-94.
DOI: https://doi.org/10.1097/RLI.0000000000000312

Noise Texture Deviation
A Measure for Quantifying Artifacts in Computed Tomography Images With Iterative Reconstructions

Fabian Morsbach, MD, Lotus Desbiolles, MD,*† Rainer Raupach, PhD,‡ Sebastian Leschka, MD,*† Bernhard Schmidt, PhD,‡ and Hatem Alkadhi, MD, MPH, EBCR**

Objectives: The aims of this study were to introduce the measure noise texture deviation as quantitative parameter for evaluating iterative reconstruction (IR)–specific artifacts in computed tomography (CT) images and to test whether IR-specific artifacts, quantified through this measure, are reduced in advanced modeled IR (ADMIRE) as compared with sinogram-affirmed IR (SAFIRE) images of the liver ex vivo and in patients with hypodense liver lesions.

Materials and Methods: In the ex vivo study part, an abdominal phantom was used. In the institutional review board–approved in vivo study part, 40 consecutive patients (mean age, 63 years) with hypodense liver lesions undergoing abdominal CT in the portal-venous phase were included. Images were reconstructed with filtered back projection, with the second-generation IR algorithm SAFIRE and with the third-generation IR algorithm ADMIRE. Noise power spectra and noise texture deviation were calculated in the phantom; image noise was measured in the phantom and in patients. Two blinded readers evaluated all image data regarding IR-specific artifacts (plastic-like, blotchy appearance); patient data were evaluated regarding conspicuity and confidence for detecting hypodense liver lesions.

Results: Image noise was significantly reduced at increasing IR levels ($P < 0.001$) with both algorithms, with no significant differences between corresponding strength levels of SAFIRE and ADMIRE (all, $P > 0.05$). Noise power spectra were similar at corresponding strength levels of SAFIRE and ADMIRE (all, $P > 0.05$). Noise texture deviation in ADMIRE was reduced compared with corresponding strength levels of SAFIRE (all, $P < 0.001$) and strongly correlated with subjective IR-specific artifacts ($r = 0.88$, $P < 0.001$). Iterative reconstruction–specific artifacts were significantly reduced in ADMIRE compared with that in SAFIRE images at strength levels 3 or greater, both ex vivo and in vivo (all, $P < 0.001$). There were no significant differences in the readers' ratings of lesion conspicuity and lesion confidence in detecting hypodense liver lesions between SAFIRE and ADMIRE ($P > 0.05$). Only lesion conspicuity was superior with SAFIRE and ADMIRE compared with filtered back projection (all, $P < 0.001$).

Conclusions: Noise texture deviation is a quantitative measure reflecting IR-specific artifacts and is reduced in CT images with ADMIRE compared with SAFIRE.

Key Words: multidetector computed tomography, liver, image processing

(*Invest Radiol* 2016;00: 00–00)

Received for publication March 24, 2016; and accepted for publication, after revision, June 21, 2016.

From the *Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich; †Clinic of Radiology and Nuclear Medicine, Kantonsspital St Gallen, St Gallen, Switzerland; and ‡Siemens, Siemens Healthcare, Forchheim, Germany.

Conflicts of interest and sources of funding: RR and BS are employees of Siemens Healthcare.

The authors declare no conflict of interest.

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.investigativeradiology.com).

Correspondence to: Hatem Alkadhi, MD, MPH, EBCR, Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Raemistrasse 100, CH-8091 Zurich, Switzerland. E-mail: hatem.alkadhi@usz.ch.

ISSN: 0020-9996/16/0000–0000

DOI: 10.1097/RLI.0000000000000312

Iterative reconstruction (IR) algorithms have been introduced for computed tomography (CT) imaging for various indications and body regions.[1] However, a plastic-like, blotchy image appearance, hereafter called *IR-specific artifact*, has been repetitively described with various IR algorithms, potentially affecting and deteriorating image quality. Such IR-specific artifacts can hamper the evaluation of images and interpretation of imaging findings.[2–8] Moreover, because this image appearance usually is pronounced at higher levels of IR algorithms,[2,6,9,10] it potentially precludes exploitation of the full potential of IR for reducing the radiation dose of the respective CT examinations.[11,12]

One recent IR technique is called *advanced modeled IR* (ADMIRE), which is the successor of sinogram-affirmed IR (SAFIRE) from the same vendor. Advanced modeled IR is characterized by improvements in the statistical modeling applied to the raw projection data including the detector signal domain. Moreover, it uses a 3-dimensional reconstruction domain with a larger neighbor volume for validation and better preservation of the CT noise texture and for artifact suppression.[11] One of the main aims of ADMIRE is to reduce the IR-specific artifacts described previously, which are known to occur at higher strength levels of the predecessor algorithm SAFIRE.[2,6,9,10] Objective evaluations of the quality of IR images in CT have been performed using the parameter noise, defined as the standard deviation of attenuation in a defined region of interest (ROI), or by calculating noise power spectra (NPS).[4,13] Previous studies used subjective image quality scores or just mentioned this artifact as a shortcoming of IR images.[2,5,10,11,14] The lack of an objective measure allowing for quantification of the IR-specific artifacts, however, precludes meaningful comparisons among different IR algorithms.

A recent in vivo study demonstrated that use of ADMIRE in abdominal CT increases image quality by lowering noise when compared with filtered back projection (FBP) and showed that the attenuation values in various organs remained constant across the different strength levels of ADMIRE.[15] A recent phantom study suggested that ADMIRE improves low-contrast detectability compared with FBP,[11] but no study, to our knowledge, has evaluated whether IR-specific artifacts in image impression are reduced with ADMIRE and whether low-contrast detectability in the liver is preserved in vivo with this type of IR algorithm as compared with SAFIRE. In addition, no objective metric was introduced so far for quantifying image artifacts specifically related to IR.

The purpose of this study was thus 2-fold: first, to introduce the measure noise texture deviation as a quantitative parameter for evaluating IR-specific artifacts in CT images; and second, to test whether these IR-specific artifacts, as quantified by this measure, are reduced in ADMIRE as compared with SAFIRE images of the liver ex vivo and in patients with hypodense liver lesions.

MATERIALS AND METHODS

Data Acquisition and Reconstruction

Phantom Study

A mathematical phantom was designed to mimic an adult abdominal cross section including the liver, spleen, osseous spine, and abdominal wall soft tissue (Fig. 1). The lateral diameter of the phantom

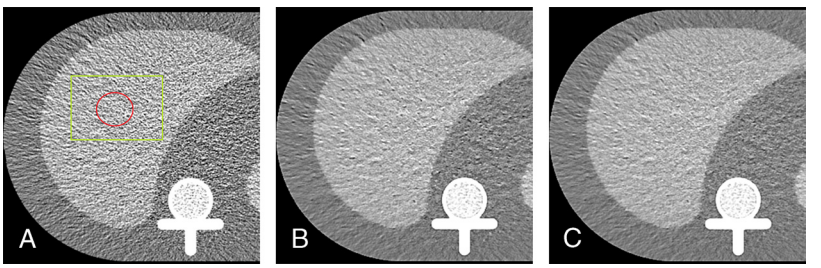

FIGURE 1. Transverse CT images of the phantom reconstructed with (A) FBP (including the ROIs placed in the liver [circular, noise; rectangular, noise texture deviation]), (B) SAFIRE strength level 5, and with (C) ADMIRE strength level 5. Figure 1 can be viewed online in color at www.investigativeradiology.com.

was 42 cm; the anteroposterior diameter was 28 cm. The simulated attenuation values for the liver, spleen, and abdominal wall soft tissue were 90, 100, and 35 HU, respectively, at 120 kVp.

We opted for a mathematical instead of real phantom for the following reason: the newly introduced metric of this study (ie, noise texture evaluation) is sensitive to even small homogeneity errors. These would deteriorate the value of results on statistical measures such as histograms. The mathematical definition of sinogram data ensures that there is no pseudostatistical influence due to inhomogeneity. The noise statistics on top of ideal noise-free CT raw data (signal-dependent Poisson noise plus electronic noise floor) and scanner geometry is well understood and relatively easy to model. We were investigating the performance of reconstruction algorithms rather than that of a certain CT scanner. Because we apply identical reconstruction algorithms as for measured data, the nontrivial parts of noise propagation from projection to image data are actually identical.

CT Data Acquisition and Reconstruction—Phantom

Using the parameters of the vendor's default abdominal portal-venous CT protocol, synthetic raw data were generated (CTSIM simulation software package; Siemens Healthcare, Forchheim, Germany) with realistic data noise that accounted both for quantum mottle and electronic noise of the detector. The geometry of the simulated CT scanner matched the real CT scanner used in the patient study (see below).

Computed tomography images were reconstructed with a slice thickness of 2 mm and increment of 1.6 mm. The field of view was set at 450 mm to include the entire abdomen; the pixel matrix was 512 × 512.

Similar to SAFIRE, the parameters and criteria of the noise model of ADMIRE can be chosen by the user to obtain a certain predefined noise reduction, and 5 presets (strength levels 1 to 5) are available for adaptation of the noise model and for controlling image impression and noise reduction. The strength level is not related to the number of iteration loops. From the raw data, the following reconstructions were made: FBP (B30f), SAFIRE at strength levels 1 to 5 (I30f), and ADMIRE at strength levels 1 to 5 (I30f).

Patient Study

In the patient study part, data from a total of 810 patients who underwent clinically indicated abdominal CT in our department between October and December 2013 were screened for inclusion. Inclusion criteria were 1 or more hypodense liver lesions of known origin and/or known etiology (eg, previous examinations including various imaging modalities and—in patients with metastases—histopathological proof of underlying colorectal cancer) and an abdominal CT acquired in the portal-venous phase of enhancement. The final study population included 40 patients (21 male, mean age 63 years, range 39–80 years; 19 female, mean age 62 years, range 24–79 years) with liver cysts (n = 25) and metastases from colorectal carcinoma (n = 15) (for the demographic data, see Table 1). Because some patients had more than 1 lesion, a total of 36 liver cysts and 19 metastases were evaluated.

The patient study part had local ethics committee approval. Written informed consent requirement was waived because of the retrospective design of this study part; all CT scans were clinically indicated, and none was performed merely for the purpose of the study.

CT Data Acquisition and Reconstruction—Patients

All CT data in patients were acquired with a 64-slice CT scanner (Somatom Definition AS 64; Siemens AG, Healthcare Sector, Forchheim, Germany). Patients received a fixed amount of 100 mL contrast media (iopromidum, Ultravist 370; 370 mg I/mL, Bayer Schering Pharma, Berlin, Germany) at a flow rate of 3 mL/s. Contrast media was injected over an antecubital vein using a dual-head power injector (Stellant; Medrad, Indianola, PA). Image acquisition was performed 70 seconds after contrast injection in the portal-venous phase. Our institutional default setting (SAFIRE strength level 3) was used as the standard reconstruction algorithm. Tube voltage and tube current was adjusted for each patient depending of the patient's weight and abdominal diameter using automatic attenuation-based tube voltage adaptation (CAREkV, slider position 7/soft tissue) and automatic

TABLE 1. Patient Demographics and CT Scan Protocol

Patient Demographics	
Total no. patients	40
Age, mean ± SD, y	62.7 ± 14.2
Sex ratio, male/female	21/19
Body weight, kg	76.1 ± 23.1
Height, cm	170.0 ± 10.0
Body mass index, kg/cm^2	26.4 ± 7.8
Diameter anteroposterior, mean ± SD, mm	249.7 ± 46.8
Diameter lateral, mean ± SD, mm	327.9 ± 36.6
No. patients with liver metastasis	15
No. patients with liver cysts	25
Scanning Parameters	
Slice collimation, mm	64 × 0.6
Effective reference tube voltage, kVp	103 ± 7 (100–120)
Effective tube current-time product, mA	197 ± 81 (110–484)
Rotation time, s	0.5
Pitch	0.9
Radiation dose	
$CTDI_{vol}$, mGy	9.3 ± 3.8
SSDE, mGy	12.1 ± 5.0

CT indicates computed tomography; SD, standard deviation; $CTDI_{vol}$, volume computed tomography dose index; SSDE, size-specific dose estimate.

attenuation-based tube current modulation (CAREDose4D). The radiation dose parameter volume CT dose index ($CTDI_{vol}$) was taken from the electronically lodged protocol from each CT study. Anterior-posterior and lateral dimensions of the patients at the level of the celiac trunk were measured for calculating size-specific dose estimates[16] (for further protocol details and radiation doses, see Table 1). Image analyses were performed on a high-definition liquid crystal display monitor (BARCO Medical Imaging Systems, Kortrijk, Belgium) using the picture archiving and communication system of our hospital (Impax Version 6.5.5.1003; Agfa-Gevaert, Mortsel, Belgium).

Raw data of the CT scans were transferred to an offline workstation equipped with a prototype software allowing for reconstruction of FBP, SAFIRE, and ADMIRE images from the same raw data using the same reconstruction kernels (see below).

All CT images were reconstructed with a slice thickness of 2 mm, increment of 1.6 mm, and a medium tissue convolution kernel (B30f for FBP as well as I30f for SAFIRE and ADMIRE, respectively). Strength levels 3 to 5 for SAFIRE and ADMIRE were chosen for reconstruction of the patient data, because level 3[5,17] is most commonly used for clinical image evaluation in SAFIRE, and levels 4 and 5 were attributed with IR-specific artifacts.[5,17] The field of view was 450 mm, as in the phantom study; the pixel matrix was 512 × 512.

Image Analysis

Objective Image Quality—Phantom

Image noise, defined as standard deviation of attenuation (in Hounsfield units, HU), was determined by 1 reader (FM [blinded for review], with 3 years of experience in abdominal radiology) who placed a circular ROI of 500 mm^2 in the liver. The liver and other simulated tissues in the phantom are homogenous; the ROIs were placed in the center of the respective organ. Signal-to-noise ratio (SNR) was calculated by dividing the mean attenuation (in Hounsfield units) by the standard deviation of attenuation (in Hounsfield units). Contrast-to-noise ratio (CNR) was calculated by dividing the difference in attenuation between liver parenchyma and abdominal wall soft tissue (in Hounsfield units) by the standard deviation of attenuation in the liver (in Hounsfield units). Noise power spectra were calculated as previously described[18] and were applied to a rectangular ROI of 2000 mm^2 in the liver. From the discrete noise power spectra NPS(f_1), the spatial frequency with maximum amplitude, $f_{max} = \max\{\mathrm{NPS}(f_i)\}$, and the mean frequency were calculated according to $f_{mean} = \sum_i f_i \times \mathrm{NPS}(f_i)/\sum_i \mathrm{NPS}(f_i)$.

Noise texture deviation was determined by analyzing the histogram of attenuation coefficients in the same rectangular ROI used previously. Let p_i be the HU values of pixels in the respective ROI, Ω. From the pixel values, the mean value $\bar{p} = \frac{1}{N_\Omega}\sum_{i\in\Omega}(p_i)(N_\Omega = |\Omega|)$ ($N_\Omega = |\Omega|$, number of pixels in the ROI Ω) and the noise standard deviation $\sigma = \sqrt{\frac{1}{N_\Omega - 1}\sum_{i\in\Omega}(p_i - \bar{p})^2}$ were calculated. Although there are numerous possibilities to $\{p_i\}$ characterize statistical distributions, we chose a simple approach of counting outlying pixels of the distribution that deviate more than 3σ from their mean value. To be independent of the ROI size, we calculated the fraction of outliers relative to the number of pixels in the ROI: $f_{3\sigma} = \frac{1}{N_\Omega}|\{i||p_i - \bar{p}| > 3\sigma\}|$. A normal distribution as expected for linear reconstruction methods such as FBP would show an outlier fraction of approximately $f_{3\sigma} = 0.27\%$ according to this metrics.[19] Three standard deviations as threshold were chosen for the following reasons. For the expected (almost) ideal normal distribution of attenuation values in FBP images, the amount of outliers is well below 1% (see above). The noise reduction of the analyzed IR techniques is at the order of up to 50%. Outliers in those images with IR can be regarded as "survivors" of extreme, but relatively rare statistical deviations of the statistical distribution in the FBP images. Outliers with 3σ at 50% noise reduction correspond to pixels exceeding only 1.5σ of the original (ie, FBP) noise level, which have a much higher probability, $f_{1.5\sigma} = 13.4\%$. Thus, relevant outliers after application of the nonlinear reconstruction techniques become visible with a 3σ threshold with respect to the new noise standard deviation level (Fig. 2).

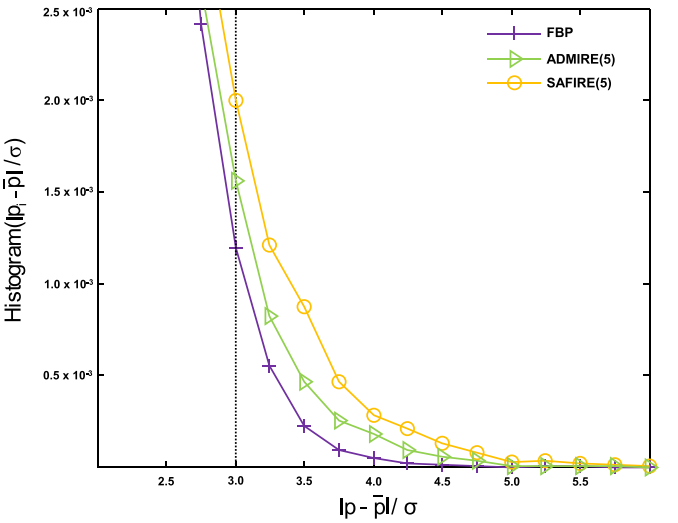

FIGURE 2. Histogram of the noise in relation to the noise deviations shown for FBP, SAFIRE (strength level 5), and ADMIRE (strength level 5). The x-axis indicates the deviation of a pixel from the mean of all pixels in an ROI normalized to the standard deviation of the noise. The y-axis indicates the histogram of all pixels in the ROI. Note that a standard deviation of 3σ (vertical dotted line) distinguishes well between the reconstruction algorithms.

Subjective Image Quality—Phantom

Subjective image quality was assessed by 2 other blinded and independent readers (LD [blinded for review] and SL [blinded for review], with 4 and 5 years of experience in abdominal radiology, respectively). Subjective image noise and IR-specific artifacts were assessed as previously shown.[2] Noise was rated subjectively on a scale from 1 (minimal image noise) to 5 (unacceptable image noise). Iterative reconstruction–specific artifacts (plastic-like, blotchy image appearance) were rated on a scale from 1 (no artifacts) to 4 (artifacts affecting diagnostic information).[2]

Objective Image Quality—Patients

Image noise was determined by 1 reader (xx [blinded for review]) who placed circular ROIs in the liver, spleen, and in the subcutaneous fat of the abdominal wall of patients. Region of interest size (mean size 436 ± 58 mm^2) was adjusted to include as much of the respective tissue without including lesions and larger vessels. The standard deviation of attenuation (in Hounsfield units) in the ROI was noted as measure of image noise, similar to the phantom measurements.

Signal-to-noise ratio and CNR were calculated similar to the phantom study.

Subjective Image Quality—Patients

Four weeks after the read-out of the phantom images, 2 readers (yy [blinded for review] and zz [blinded for review]) assessed the subjective image quality in patients. The readers were blinded to the reconstruction algorithm and patient data. Image noise and IR-specific artifacts were assessed according to the same criteria mentioned previously. In addition, hypodense liver lesions were assessed in regard to their conspicuity ranging on a scale from 1 (well-seen lesion with well-visualized margins) to 5 (definitely an artifact mimicking a lesion), and diagnostic confidence ranging on a scale from 1 (completely confident) to 3 (poor confidence).[2]

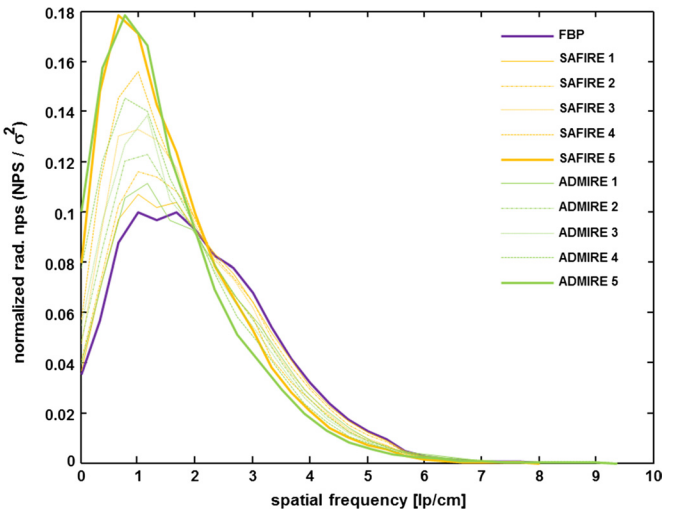

FIGURE 3. Noise power spectra plots for FBP, SAFIRE, and ADMIRE. Note that the plots do not distinguish between corresponding strength levels of SAFIRE and ADMIRE.

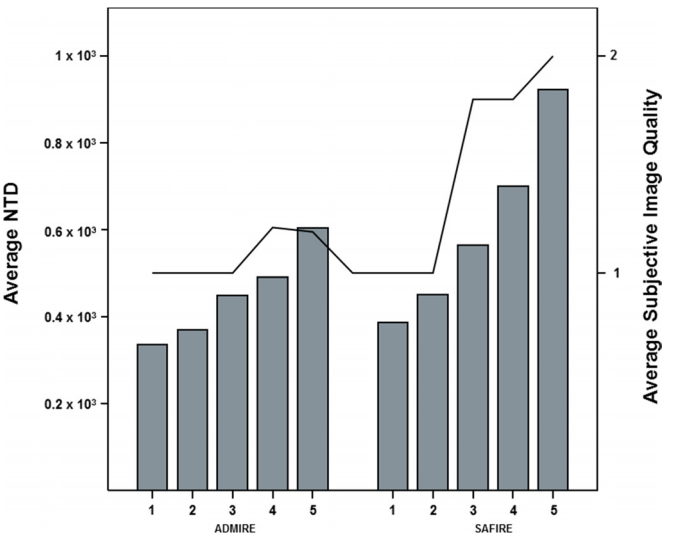

FIGURE 4. Correlation of the increase in IR-specific artifacts and noise texture deviation (NTD) at increasing strengths levels. Boxes represent NTD; line represents IR-specific artifact scores. Figure 4 can be viewed online in color at www.investigativeradiology.com.

Statistical Analysis

Variables were expressed as mean ± standard deviation. Descriptive variables were expressed as frequencies. The Shapiro-Wilk test was used to test for normality.

The interreader agreement regarding qualitative evaluation of the readers was analyzed by using the intraclass correlation coefficient.[20] Agreement was interpreted as follows: poor, less than 0.20; fair, 0.21–0.40; moderate, 0.41–0.60; good, 0.61–0.80; and very good, 0.81–1.00.[21]

Because of paired data, Friedman test was performed for evaluating significant differences in objective and subjective image quality between data sets. Post hoc testing was performed using the Wilcoxon signed rank test. Spearman ρ was used to test for a correlation between noise texture deviation and IR-specific artifact scores and between noise texture deviation and ADMIRE and SAFIRE strength levels, respectively, as well for a correlation between f_{mean} and f_{max} values from the NPS analysis and IR-specific artifact scores.

A 2-tailed P value below 0.05 indicated statistical significance. Post hoc tests were considered to be statistically significant for P values of less than 0.007 in concordance with Bonferroni adjustment for multiple comparisons. All statistical analyses were conducted using commercially available software (SPSS, release 21.0; SPSS, Chicago, IL).

RESULTS

Image Quality—Phantom

Image noise was significantly different between FBP and SAFIRE 1–5 ($P < 0.001$), and between FBP and ADMIRE 1–5 ($P < 0.001$) and significantly decreased with each higher strength level of ADMIRE and SAFIRE (all, $P < 0.001$). No significant differences in noise in the liver, spleen and soft tissue between corresponding strength levels of SAFIRE and ADMIRE were found (all, $P > 0.05$).

Signal-to-noise ratio and CNR was significantly different between FBP and SAFIRE 1–5 ($P < 0.001$), and between FBP and ADMIRE 1–5 ($P < 0.001$) and significantly increased with each higher strength level of ADMIRE and SAFIRE (all, $P < 0.001$). No significant differences in SNR and CNR between corresponding strength levels of SAFIRE and ADMIRE were found (all, $P > 0.05$).

Noise power spectra showed no significant differences between data sets for f_{max} and f_{mean} (both, $P > 0.05$) (Fig. 3).

Noise texture deviation showed significant overall differences between reconstruction algorithms ($P < 0.001$). Increasing strength levels of ADMIRE and SAFIRE correlated with an increase in noise texture deviation ($r = 0.95$, $P < 0.001$). There were significant pairwise differences in noise texture deviations between SAFIRE 1 and ADMIRE 1 (4×10^{-3} vs 3×10^{-3}, $P < 0.001$), SAFIRE 2 and ADMIRE 2 (5×10^{-3} vs 4×10^{-3}, $P < 0.001$), SAFIRE 3 and ADMIRE 3 (6×10^{-3} vs 5×10^{-3}, $P < 0.001$), SAFIRE 4 and ADMIRE 4 (7×10^{-3} vs 5×10^{-3}, $P < 0.001$), and between SAFIRE 5 and ADMIRE 5 (9×10^{-3} vs 6×10^{-3}, $P < 0.001$) (Table 2).

Subjective image noise progressively decreased from FBP to increasing strength levels of SAFIRE 1–5 (scores 4 vs 4 vs 3 vs 3 vs 2; all, $P < 0.001$) and ADMIRE (scores 4 vs 4 vs 3 vs 3 vs 2; all, $P < 0.001$). There were no significant differences in subjective noise between corresponding strength levels of SAFIRE and ADMIRE (level 1: score 4 vs 4, $P = 0.88$; level 2: score 4 vs 4, $P = 0.11$; level 3: score 3 vs 3, $P = 0.78$; level 4: score 3 vs 3, $P = 0.11$; level 5: score 2 vs 2, $P = 0.08$).

TABLE 2. Mean Values of the Quantitative Parameter Noise Texture Deviation in the Phantom Liver

Strength Level		1	2	3	4	5
FBP	3.4×10^{-3}					
SAFIRE		3.9×10^{-3}	4.5×10^{-3}	5.6×10^{-3}	7.0×10^{-3}	9.2×10^{-3}
ADMIRE		3.4×10^{-3}	3.7×10^{-3}	4.5×10^{-3}	4.9×10^{-3}	6.0×10^{-3}
P		<0.001	<0.001	<0.001	<0.001	<0.001

FBP indicates filtered back projection; SAFIRE, sinogram-affirmed iterative reconstruction; ADMIRE, advanced modeled iterative reconstruction.

Iterative reconstruction–specific artifacts were significantly less on ADMIRE than on corresponding SAFIRE images at strength levels 3 or greater (level 1: score 1 vs 1, $P = 0.07$; level 2: score 1 vs 1, $P = 0.08$; level 3: score 2 vs 1, $P < 0.001$; level 4: score 2 vs 1, $P < 0.001$; level 5: score 2 vs 1, $P < 0.001$). Iterative reconstruction–specific artifacts in images with both IR algorithms strongly correlated with noise texture deviation ($r = 0.88$, $P < 0.001$) (Fig. 4). No significant correlation was found between IR-specific artifacts and the parameters noise ($P > 0.05$) and the f_{mean} and f_{max} from NPS analysis ($P > 0.05$).

Image Quality—Patients

Image noise was significantly different between FBP and SAFIRE levels 1–5 ($P < 0.001$) and between FBP and ADMIRE level 1–5 ($P < 0.001$, Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/RLI/A299).

There was an average noise reduction of 33% comparing ADMIRE and SAFIRE level 3 to FBP. Further noise reduction from level 3 to 4 was on average 13% for SAFIRE and 14% for ADMIRE, and from level 4 to 5 on average 15% for SAFIRE and 17% for ADMIRE. No significant differences in image noise were found between corresponding SAFIRE and ADMIRE strength levels (all $P > 0.05$, Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/RLI/A299).

Signal-to-noise ratio and CNR was significantly different between FBP and SAFIRE 1–5 ($P < 0.001$), and between FBP and ADMIRE 1–5 ($P < 0.001$) and significantly increased with each higher strength level of ADMIRE and SAFIRE (all, $P < 0.001$). No significant differences in SNR and CNR between corresponding strength levels of SAFIRE and ADMIRE were found (all, $P > 0.05$).

The interreader agreement for subjective image quality was good (intraclass correlation coefficient, 0.84; $P < 0.001$). Thus, the readings from 1 reader (yy [blinded for review]) were used for further analyses.

Subjective Image Noise

There were significant differences in subjective image noise among data sets ($P < 0.001$). We found significant differences between

TABLE 3. Frequency of Image Quality Scores and Lesion Conspicuity/Confidence for Detection of Hypodense Liver Lesions in Patients on Images With FBP, SAFIRE, and ADMIRE at Strength Levels 3–5

	FBP	SAFIRE 3	SAFIRE 4	SAFIRE 5	ADMIRE 3	ADMIRE 4	ADMIRE 5
Subjective image noise							
Score 1	0/40	0/40	5/40	12/40	0/40	5/40	12/40
Score 2	2/40	8/40	12/40	22/40	13/40	32/40	23/40
Score 3	6/40	30/40	23/40	6/40	25/40	3/40	5/40
Score 4	32/40	2/40	0/40	0/40	2/40	0/40	0/40
Score 5	0/40	0/40	0/40	0/40	0/40	0/40	0/40
IR-specific artifacts							
Score 1	NA	6/40	0/40	0/40	28/40	22/40	20/40
Score 2	NA	32/40	35/40	29/40	12/40	17/40	17/40
Score 3	NA	2/40	5/40	11/40	0/40	1/40	3/40
Score 4	NA	0/40	0/40	0/40	0/40	0/40	0/40
Lesion evaluation							
Metastasis							
Lesion conspicuity							
Score 1	10/19	12/19	18/19	19/19	15/19	18/19	19/19
Score 2	7/19	6/19	1/19	0/19	4/19	1/19	0/19
Score 3	2/19	1/19	0/19	0/19	0/19	0/19	0/19
Score 4	0/19	0/19	0/19	0/19	0/19	0/19	0/19
Score 5	0/19	0/19	0/19	0/19	0/19	0/19	0/19
Diagnostic confidence							
Score 1	18/19	18/19	19/19	19/19	19/19	19/19	19/19
Score 2	1/19	1/19	0/19	0/19	0/19	0/19	0/19
Score 3	0/19	0/19	0/19	0/19	0/19	0/19	0/19
Cysts							
Lesion conspicuity							
Score 1	16/36	25/36	34/36	35/36	33/36	36/36	36/36
Score 2	20/36	10/36	1/36	0/36	3/36	0/36	0/36
Score 3	0/36	1/36	1/36	1/36	0/36	0/36	0/36
Score 4	0/36	0/36	0/36	0/36	0/36	0/36	0/36
Score 5	0/36	0/36	0/36	0/36	0/36	0/36	0/36
Diagnostic confidence							
Score 1	36/36	36/36	36/36	36/36	36/36	36/36	36/36
Score 2	0/36	0/36	0/36	0/36	0/36	0/36	0/36
Score 3	0/36	0/36	0/36	0/36	0/36	0/36	0/36

FBP indicates filtered back projection; SAFIRE, sinogram-affirmed iterative reconstruction; ADMIRE, advanced modeled iterative reconstruction; IR, iterative reconstruction; NA, not applicable.

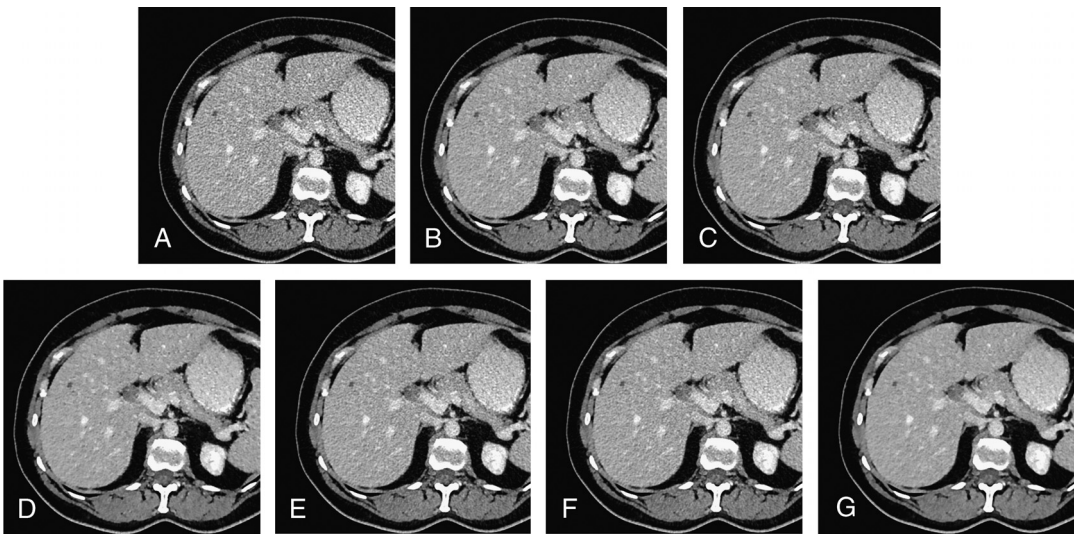

FIGURE 5. Transverse CT images in a 47-year-old female patient with a small cyst in liver segment V reconstructed with FBP (A), SAFIRE strength level 3 (B), 4 (C), 5 (D), and ADMIRE strength level 3 (E), 4 (F), 5 (G). Note the reduced IR-specific artifacts (pixilated, blotchy image appearance) on ADMIRE as compared with SAFIRE images, especially for ADMIRE 5.

FBP and SAFIRE levels 3–5 (scores 4 vs 2 vs 2 vs 1, $P < 0.001$) and between FBP and ADMIRE levels 3–5 (scores 4 vs 2 vs 2 vs 1, $P < 0.001$). No significant differences in subjective image noise were found between corresponding levels of SAFIRE and ADMIRE (all $P > 0.05$, Table 3).

IR-Specific Artifacts

There were significant overall differences in IR-specific artifacts among data sets ($P < 0.001$) (Figs. 5, 6), with significantly less such artifacts in ADMIRE compared with corresponding SAFIRE levels (level 3: score 2 vs 1, $P < 0.001$; level 4: score 2 vs 1, $P < 0.001$; level 5: score 2 vs 1, $P < 0.001$, Table 3).

Conspicuity

There were significant overall differences in regard to the conspicuity of metastases ($P < 0.001$) and cysts ($P < 0.001$) between reconstruction algorithms. We found significant differences between FBP and SAFIRE levels 3–5 (scores 2 vs 1 vs 1 vs 1, $P < 0.001$) and between FBP and ADMIRE levels 3–5 (scores 2 vs 1 vs 1 vs 1, $P < 0.001$). No significant differences in conspicuity of metastases and cysts between corresponding strength levels of SAFIRE and ADMIRE were found (all, $P > 0.05$, Table 3).

Diagnostic Confidence

There were no significant overall differences in regard to the confidence for detecting metastases and cysts between reconstruction algorithms (all, $P > 0.05$, Table 3).

DISCUSSION

Previous studies used subjective image quality scores for describing the plastic-like, blotchy image appearance that often is encountered in CT images reconstructed at higher levels of IR.[2,5,10,11,14] Our study introduced noise texture deviation, being an objective measure for quantifying such IR-specific artifacts. In contrast to image noise and NPS analyses, the parameter noise texture deviation significantly

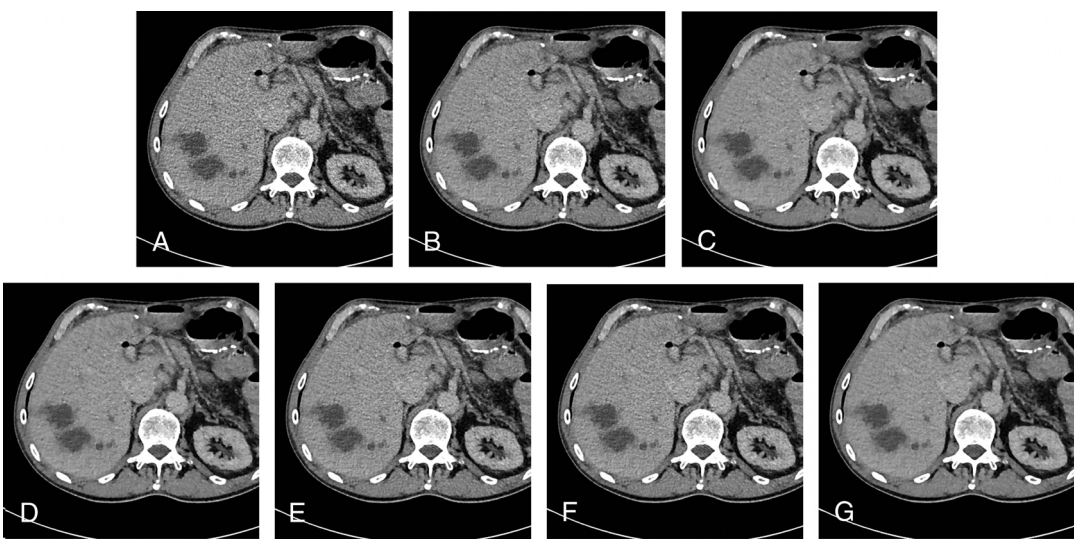

FIGURE 6. Transverse CT images in a 66-year-old male patient with liver metastases from colorectal cancer reconstructed with FBP (A), SAFIRE strength level 3 (B), 4 (C), 5 (D), and ADMIRE strength level 3 (E), 4 (F), 5 (G). Note the reduced IR-specific artifacts on ADMIRE as compared with SAFIRE images.

correlated with IR-specific artifacts. In addition, noise texture deviation was significantly reduced in CT images reconstructed with ADMIRE as compared with those reconstructed with the predecessor SAFIRE.

Measurements of the standard deviation of attenuation represent the most commonly used way for quantifying image noise in CT.[4,6,12,15,22] For the estimation of a general noise level such a measurement can be considered adequate and has shown value in optimizing radiation dose with IR.[23–25] Using this parameter for quantifying the effect of IR on image appearance, however, falls short because IR assumes a smooth image reality in the statistical model and compares neighboring pixels, whereas FBP assumes exact data and can amplify noise, which can then be reflected as the standard deviation of attenuation.[10,14,17]

In a recent study, Solomon et al[11] showed that the conventional metrics image noise and NPS are no adequate measures for evaluating the performance of CT images with IR. In our study, we found similar results at corresponding strength levels of ADMIRE and SAFIRE for both noise and NPS, being at the range of those previously reported with other IR techniques.[15,26] In addition, noise and NPS showed no correlation with IR-specific artifacts encountered in both our phantom and human liver CT data. This underlines that these 2 measures are not reflective of the plastic-like, blotchy CT image appearance encountered with IR.

To overcome the limitations of the aforementioned parameters, and for finding a way to quantitatively measure and compare IR-specific artifacts, we introduced the parameter noise texture deviation, which is based on histogram analyses of attenuation coefficients. Noise texture deviation increased with increasing strength levels of the respective IR reconstruction algorithm. We found that noise texture deviation was lower in ADMIRE images as compared with SAFIRE at all corresponding strength levels, which correlated strongly with the subjective impression of IR-specific artifacts. Given the fact that noise texture may also affect the detection performance of lesions,[7,13,27,28] the availability of an objective and quantitative measure for the image texture appears important. Furthermore, no other study yet validated whether ADMIRE actually reduces IR-specific artifacts, as announced by the vendor.

We found an average value of 3.4×10^{-3} in noise texture deviation for the liver in our phantom. This differs from the 2.7×10^{-3} we would expect for the whole volume of a water phantom in an environment with absent electronic noise. This is caused by the realistic simulation in our phantom where electronic noise plays a relevant role, particularly in the liver due to contribution of x-rays with highest attenuation in the cross section. Electronic noise in combination with taking the logarithm as input for FBP translates statistical effects into a systematic shift of mean signals toward artificially higher values. The severity of this phenomenon depends on attenuation and, therefore, on the projection angle in a rotationally nonsymmetric phantom. This effectively results in a systematic change of the statistical distribution in CT images toward hyper-Gaussian already for FBP and explains the slightly higher noise texture deviation value of 3.4×10^{-3}.

We evaluated the conspicuity and confidence for detecting hypodense liver lesions in CT images reconstructed with FBP, SAFIRE, and ADMIRE. Here we found that the conspicuity of liver lesions was improved for both SAFIRE and ADMIRE, as compared with FBP, and was similar among the 2 IR algorithms. This is most probably explained by the reduced noise and higher sharpness of IR, which is indicated by the shift toward lower spatial frequencies in the noise spectra analysis. Interestingly, the confidence for identifying hypodense liver lesions with both IR algorithms remained similar compared with FBP. This indicates that the effect on noise, image sharpness, and noise texture of both IR algorithms do not necessarily translate into an altered confidence of the readers.

The following study limitations must be acknowledged. First, noise texture deviation quantification requires that the phantom is homogeneous at least in the position of the ROI. Second, our cohort of 40 patients was relatively small. Third, we used our institutional standard abdomen CT protocol and not an explicit low-dose protocol. However, the average $CTDI_{vol}$ in our study was 9.3 mGy, which is comparable to levels previously called *low dose*.[2,26] In addition, further metrics to measure noise, such as noise measurements in subtraction images, could have been used. However, we did not find any significant differences for the metrics used in this paper and thus deemed the applied metrics sufficient. Finally, we applied the parameter noise texture deviation to IR images of 1 vendor and 1 reconstruction kernel only, and it remains to be determined whether this measure applies to other IR algorithms and kernels as well.

In conclusion, the newly proposed parameter noise texture deviation allows for quantifying IR-specific artifacts in CT image data. As opposed to image noise and NPS, noise texture deviation better correlates with the plastic-like, blotchy image appearance, which is often inherent to CT images with IR. We could show that IR-specific artifacts are significantly reduced in ADMIRE as compared with SAFIRE images, while the conspicuity and confidence for the detection of hypodense liver lesions remained preserved. The improved image impression of ADMIRE at high strength levels allows for a better exploitation of the radiation dose reduction capabilities of IR.

REFERENCES

1. Attenberger UI, Morelli J, Budjan J, et al. Fifty years of technological innovation: potential and limitations of current technologies in abdominal magnetic resonance imaging and computed tomography. *Invest Radiol*. 2015;50:584–593.
2. Singh S, Kalra MK, Hsieh J, et al. Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology*. 2010;257:373–383.
3. Kim SH, Yoon JH, Lee JH, et al. Low-dose CT for patients with clinically suspected acute appendicitis: optimal strength of sinogram affirmed iterative reconstruction for image quality and diagnostic performance. *Acta Radiol*. 2015; 56:899–907.
4. Ehman EC, Yu L, Manduca A, et al. Methods for clinical evaluation of noise reduction techniques in abdominopelvic CT. *Radiographics*. 2014;34:849–862.
5. Hardie AD, Nelson RM, Egbert R, et al. What is the preferred strength setting of the sinogram-affirmed iterative reconstruction algorithm in abdominal CT imaging? *Radiol Phys Technol*. 2015;8:60–63.
6. Fletcher JG, Krueger WR, Hough DM, et al. Pilot study of detection, radiologist confidence and image quality with sinogram-affirmed iterative reconstruction at half-routine dose level. *J Comput Assist Tomogr*. 2013;37:203–211.
7. Solomon J, Wilson J, Samei E. Characteristic image quality of a third generation dual-source MDCT scanner: noise, resolution, and detectability. *Med Phys*. 2015; 42:4941–4953.
8. Ellmann S, Kammerer F, Brand M, et al. A novel pairwise comparison-based method to determine radiation dose reduction potentials of iterative reconstruction algorithms, exemplified through circle of Willis computed tomography angiography. *Invest Radiol*. 2016;51:331–339.
9. Chang W, Lee JM, Lee K, et al. Assessment of a model-based, iterative reconstruction algorithm (MBIR) regarding image quality and dose reduction in liver computed tomography. *Invest Radiol*. 2013;48:598–606.
10. Xu J, Mahesh M, Tsui BM. Is iterative reconstruction ready for MDCT? *J Am Coll Radiol*. 2009;6:274–276.
11. Solomon J, Mileto A, Ramirez-Giraldo JC, et al. Diagnostic performance of an advanced modeled iterative reconstruction algorithm for low-contrast detectability with a third-generation dual-source multidetector CT scanner: potential for radiation dose reduction in a multireader study. *Radiology*. 2015;275:735–745.
12. Gordic S, Morsbach F, Schmidt B, et al. Ultralow-dose chest computed tomography for pulmonary nodule detection: first performance evaluation of single energy scanning with spectral shaping. *Invest Radiol*. 2014;49:465–473.
13. Siewerdsen JH, Cunningham IA, Jaffray DA. A framework for noise-power spectrum analysis of multidimensional images. *Med Phys*. 2002;29:2655–2671.
14. Fleischmann D, Boas FE. Computed tomography—old ideas and new technology. *Eur Radiol*. 2011;21:510–517.
15. Gordic S, Desbiolles L, Stolzmann P, et al. Advanced modelled iterative reconstruction for abdominal CT: qualitative and quantitative evaluation. *Clin Radiol*. 2014;69:e497–e504.

16. 204 ATG Size specific dose estimates (SSDE) in pediatric and adult CT examinations. 2011. Available at: http://www.aapm.org/pubs/reports/rpt_204.pdf. Accessed April 2015.
17. Yang WJ, Yan FH, Liu B, et al. Can sinogram-affirmed iterative (SAFIRE) reconstruction improve imaging quality on low-dose lung CT screening compared with traditional filtered back projection (FBP) reconstruction? *J Comput Assist Tomogr*. 2013;37:301–305.
18. Stierstorfer K, Spahn M. Self-normalizing method to measure the detective quantum efficiency of a wide range of x-ray detectors. *Med Phys*. 1999;26:1312–1319.
19. Bronstein IN, Semenjajew KA. *Handbook of Mathematics. Chapter 1*. 3rd revised ed. Berlin, Germany: Springer; 1985.
20. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
22. Menzel HGSH, Teunen D. *European Guidelines on Quality Criteria for Computed Tomography*. Luxembourg: European Commission; 2000. Publication No. EUR 16262 EN. 2000.
23. Prakash P, Kalra MK, Digumarthy SR, et al. Radiation dose reduction with chest computed tomography using adaptive statistical iterative reconstruction technique: initial experience. *J Comput Assist Tomogr*. 2010;34:40–45.
24. Park M, Chung YE, Lee HS, et al. Intraindividual comparison of diagnostic performance in patients with hepatic metastasis of full-dose standard and half-dose iterative reconstructions with dual-source abdominal computed tomography. *Invest Radiol*. 2014;49:195–200.
25. Vardhanabhuti V, Riordan RD, Mitchell GR, et al. Image comparative assessment using iterative reconstructions: clinical comparison of low-dose abdominal/pelvic computed tomography between adaptive statistical, model-based iterative reconstructions and traditional filtered back projection in 65 patients. *Invest Radiol*. 2014;49:209–216.
26. Schindera ST, Odedra D, Raza SA, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*. 2013;269:511–518.
27. Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am A Opt Image Sci Vis*. 1997;14:2420–2442.
28. Solomon JB, Christianson O, Samei E. Quantitative comparison of noise texture across CT scanners from different manufacturers. *Med Phys*. 2012;39:6048–6055.